

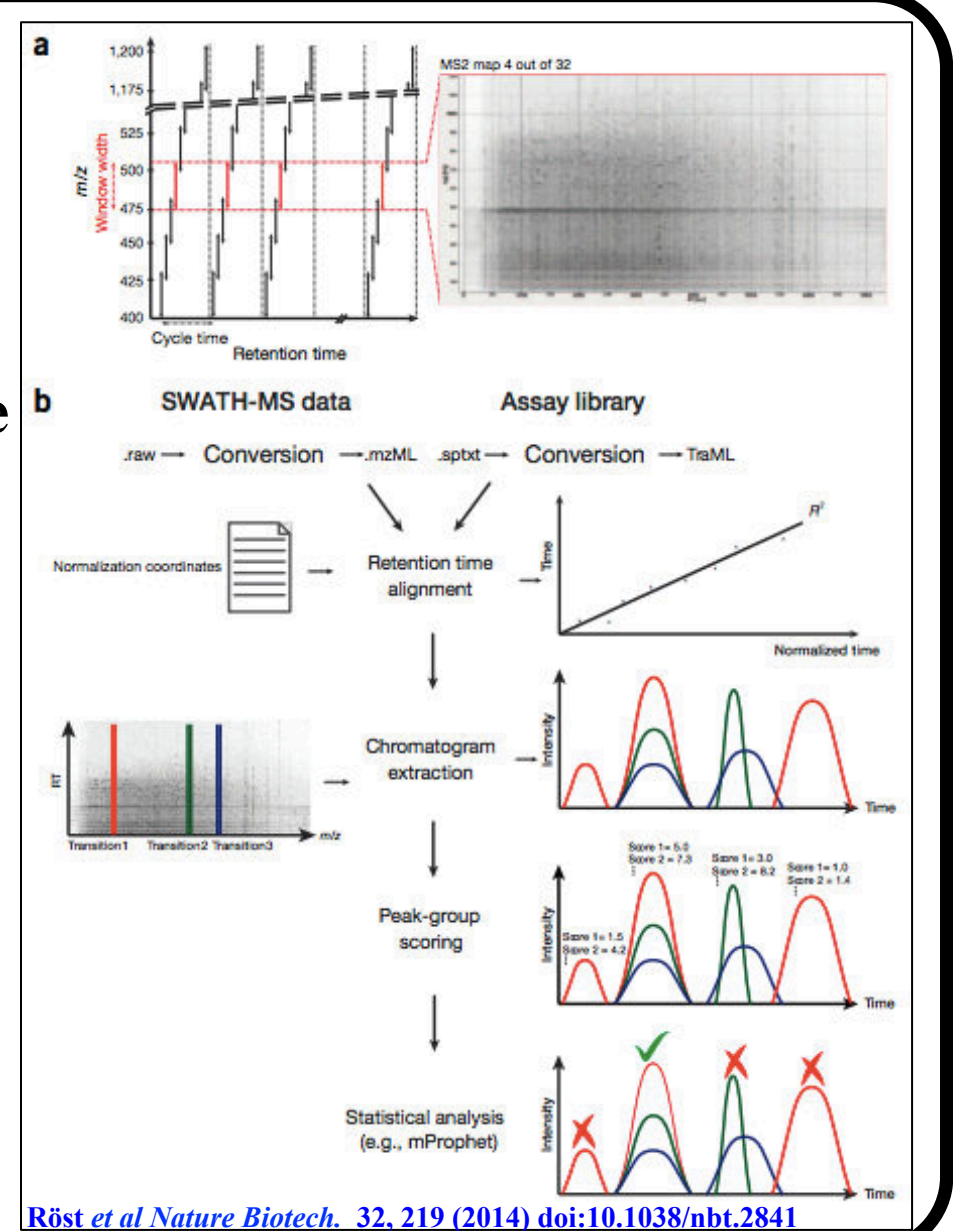
# DEMOCRATIZING AND EXPANDING THE REACH OF DIA MASS SPECTROMETRY: DEVELOPING OPENSWATH TOOLS AND WORKFLOWS WITHIN USER-FRIENDLY GALAXY-P PLATFORM

Pratik Jagtap<sup>1,2</sup>; Sarah Parker<sup>3</sup>; Bjoern Gruening<sup>4</sup>; Ira Cooke<sup>5</sup>; James Johnson<sup>6</sup>, Hannes Roest<sup>7</sup>, George Rosenberger<sup>7</sup>, Tzu-Yi Yang<sup>2</sup>, Candace Guerrero<sup>2</sup>, Brett Noel<sup>2</sup>; Laurie Parker<sup>2</sup>, Jennifer Van Eyk<sup>3</sup>, Ruedi Aebersold<sup>7,8</sup> and Tim Griffin<sup>1,2</sup>

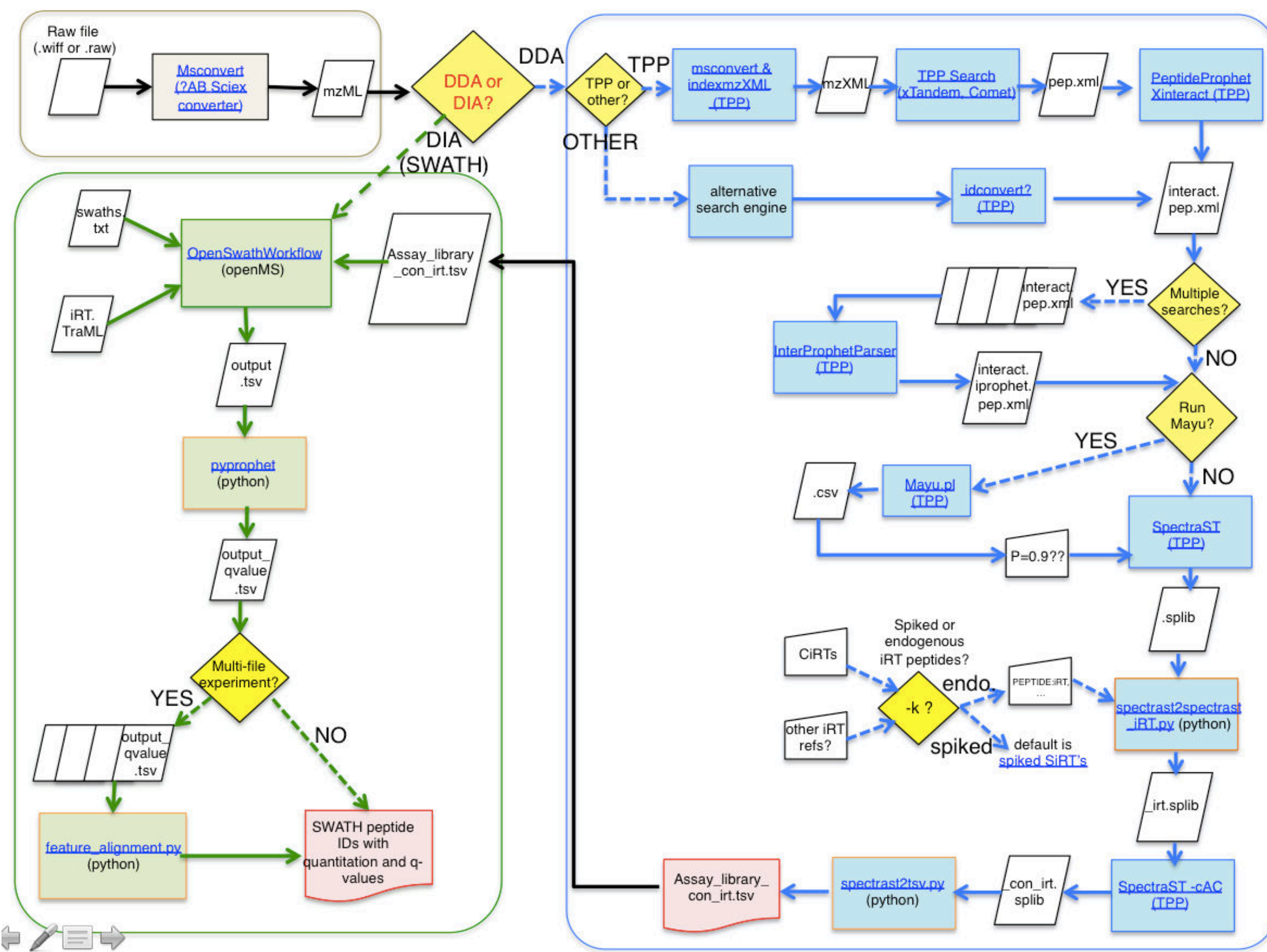
<sup>1</sup> Center for Mass Spectrometry and Proteomics, UMN, St. Paul, MN; <sup>2</sup> Biochem. Mol Biol. Biophysics, University of Minnesota, Minneapolis, MN; <sup>3</sup> Advanced Clinical Biosystems Research Institute, Los Angeles, CA; <sup>4</sup> University of Freiburg, Freiburg, Germany; <sup>5</sup> La Trobe University, Melbourne, Australia; <sup>6</sup> Minnesota Supercomputing Institute, UMN, Minneapolis, MN; <sup>7</sup> Department of Biology, IMSB, ETH Zurich, Zurich, Switzerland; <sup>8</sup> Faculty of Science, University of Zurich, Zurich, Switzerland

## OPENSWATH

- Data independent acquisition (DIA) combines the advantages of shotgun proteomics and selected reaction monitoring approaches; which generate continuous, highly complex ion fragmentation data.
- Researchers can analyze the data immediately, or archive it for a hypothesis-driven re-analysis later.
- OpenSWATH software addresses challenges of DIA data analysis, offering transparency, vendor neutrality and usable output.



## TOOLS AND WORKFLOWS SCHEMA FOR IMPLEMENTATION WITHIN GALAXYP



## OPENSWATH ANALYSIS

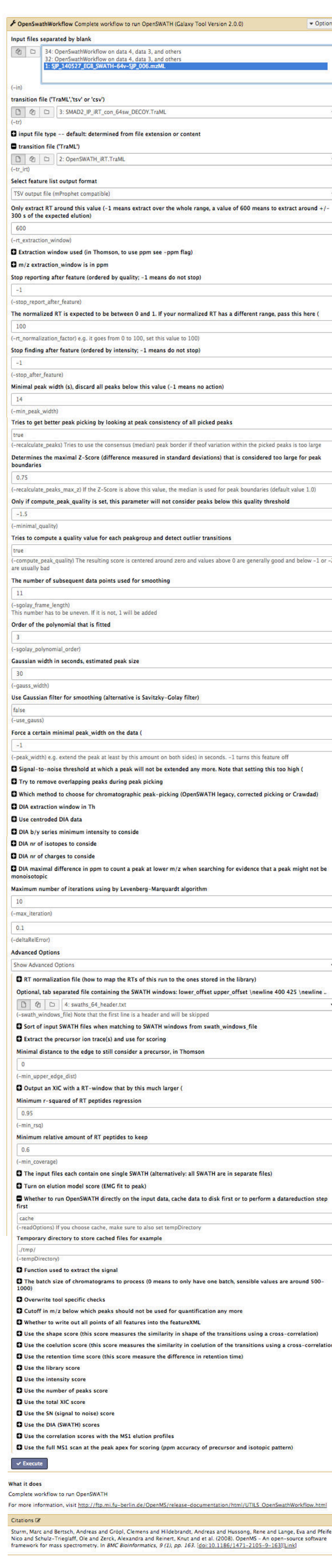
All necessary tools for OpenSWATH analysis of DIA data were wrapped into a single consolidated workflow, enabling users to:

- 1) select peptides of interest;
- 2) extract, normalize and quantify intensities for associated fragments from the data library, using software for alignment of peptides between different samples in the same experimental dataset;
- 3) statistically analyze outputs, using the pyprophet algorithm for FDR estimation etc.

## SPECTRAL LIBRARY GENERATION

MS data libraries are built from high confidence PSMs (1% FDR) via the SpectraST tool, deployed in Galaxy-P. The 'spectra2spectrast iRT' tool will be deployed for aligning retention time coordinates. After generation of a consensus MS spectrum for each identified peptide sequence, the tab separated data is converted to a .TraML file, including decoy transition groups for FDR estimation, using OpenMS tools ConvertTSVtoTraML and OpenSwathDecoyGenerator.

## OPENSWATH TOOL



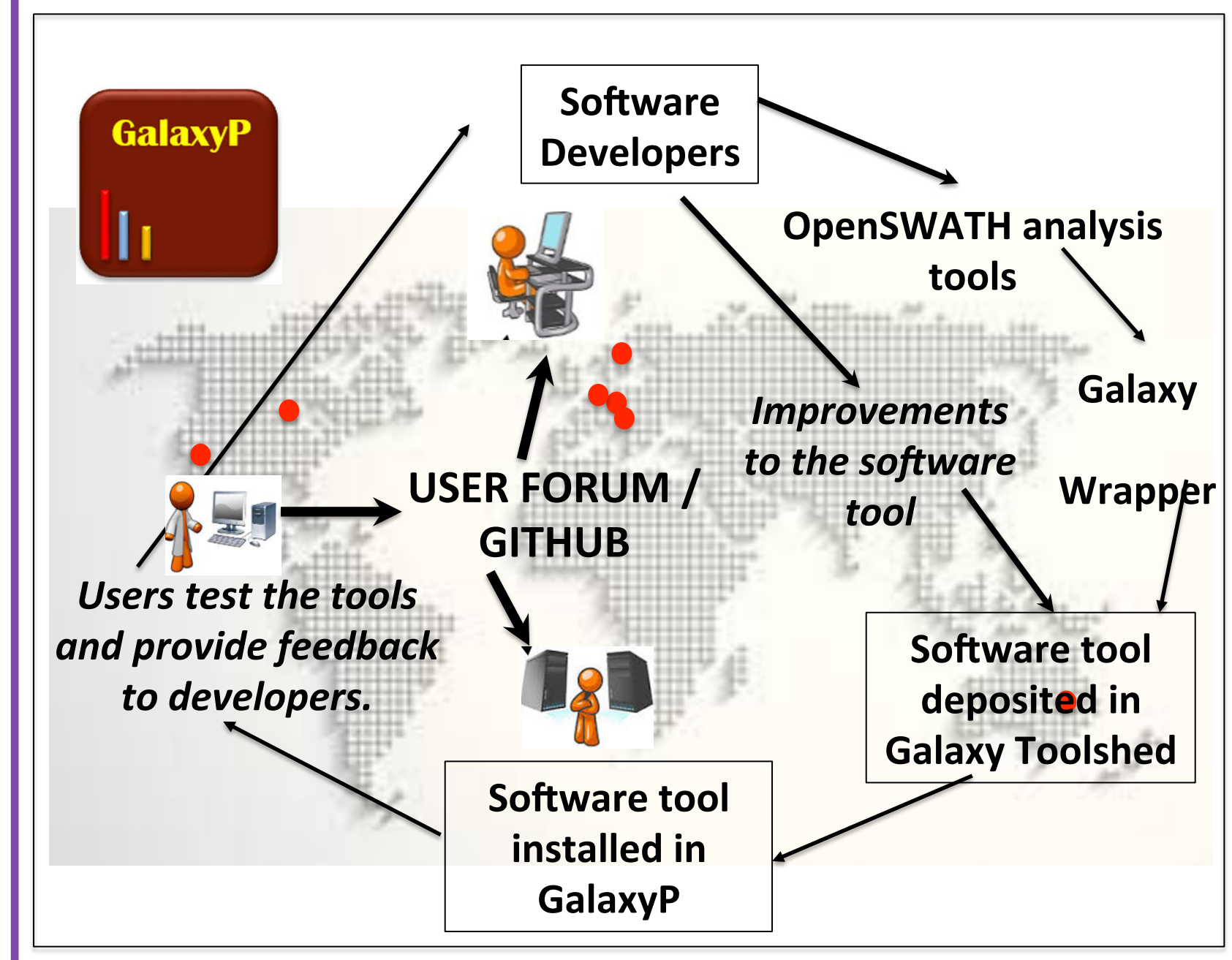
## TOOL IMPLEMENTATION STATUS

	WRAP	INSTALL	TEST
<b>DATA CONVERSION</b>			
Define .wiff datasets	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
msconvert to convert .wiff files to mzml	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
MS data Converter from AB Sciex	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>SEARCH</b>			
msconvert (ProteoWizard)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
index2XML (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
XI tandem	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Comet	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
PeptideProphet Xinteract (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
idconvert (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
InterProphetParser (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
PeptideShaker	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>LIBRARY BUILD</b>			
Mayu.pl (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SpectraST (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SpectraST2spectrast_iRT.py	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>LIBRARY REFINEMENT</b>			
SpectraST (TPP)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
SpectraST2tsv.py	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
TSVtoTraML	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
OpenSWATHDecoyGenerator	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>OPENSWATH WORKFLOW</b>			
Main OpenSWATH_Workflow (OpenMS)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
pyprophet	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Feature aligner	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<b>STATISTICAL ANALYSIS</b>			
R_MSSStats_Scripts	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

## CURRENT STATUS

- Tools such as wiff file uploader, msconvert (for wiff files), Sciex MS Data Converter, OpenSWATH Workflow and Pyprophet have been installed and tested.
- Tools within Transproteomic Pipeline outputs such as SearchGUI / PeptideShaker will also be developed and tested for integration into OpenSWATH workflow.
- The following tools have been wrapped and are on the galaxy toolshed and or github:
  - SpectraST to TSV.
  - Peptide Prophet, Interprophet and Protein Prophet from the TPP.
  - OpenMS tools.
  - Feature Aligner.
  - Pyprophet.

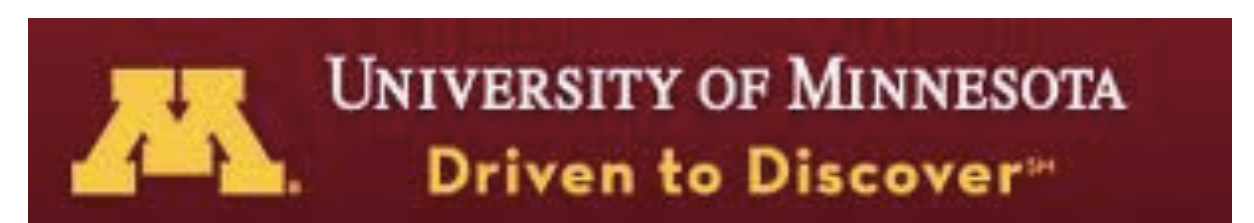
## COMMUNITY-BASED TOOL DEVELOPMENT & TESTING



Github Link: <https://github.com/galaxyproteomics>  
 Galaxy toolshed: <https://toolshed.g2.bx.psu.edu/repository>  
 GalaxyP google group: [z.umn.edu/galaxypuserforum](https://groups.google.com/g/galaxypuserforum)

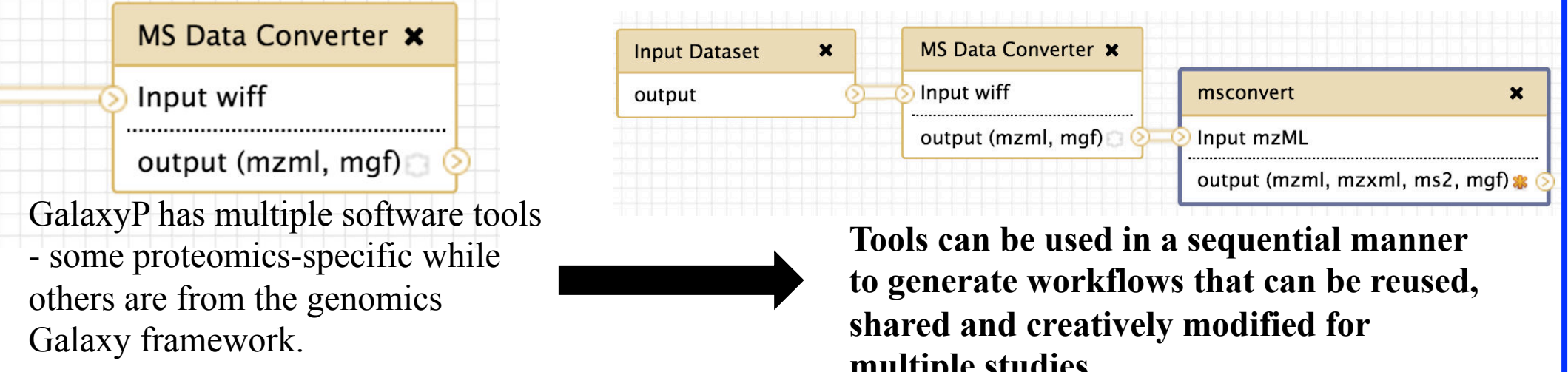
## FUTURE PLANS

- We plan to continue on our implementation of OpenSWATH DIA tools in user-friendly Galaxy framework to increase usability, dissemination and enable integration with multi-omic approaches.
- We plan to develop workflows that will offer user-friendly conversion and analysis of file formats.
- Tools that offer an alternative to Transproteomic Pipeline outputs such as SearchGUI / PeptideShaker will also be developed and tested for integration into OpenSWATH workflow.
- Downstream tools such as MSSstats for statistical quantitative analysis of selected peptides will also be integrated and tested.
- Alternative DIA workflows such as DIA-Umpire that use similar tools will also be explored.



GalaxyP is supported through the NSF Grant # 1147079.

## GalaxyP Tools



- GalaxyP has multiple software tools - some proteomics-specific while others are from the genomics Galaxy framework.
- Tools can be used in a sequential manner to generate workflows that can be reused, shared and creatively modified for multiple studies.
- Benefits of Galaxy / GalaxyP:**
- Software accessibility and usability.
  - Share-ability of tools, workflows and histories.
  - Reproducibility and ability to test and compare results after using multiple parameters.
  - Analytical transparency
  - Scalability of data

